A Temporal Knowledge Graph Dataset for Profiling

†Siraj Munir, ††Syed Imran Jami, †††Shaukat Wasi

†, ††, ††† Faculty of Computer Science, Mohammad Ali Jinnah University, Karachi, Pakistan

Summary
The Semantic representation plays a vital role in information extraction and information retrieval. Current state-of-the-art for semantic representation are Temporal Knowledge Graphs. Due to tremendous growth of things like Big Data we need a way to devise how we should represent available knowledge such that it can be more effective. Finding out semantically connected things from huge data lake is complex task which comes under the umbrella of Computer Science specifically semantic web. In this document we have presented a novel temporal Knowledge Graph dataset. Proposed dataset encompasses ten different types of relationship among which two relations are temporal based. Dataset incorporate temporal data of hundred people. For data acquisition we have used university campus environment. By using facial recognition technique backed with Convolutional Neural Network (CNN) [1] we have collected data of each person. For training our model on facial features we have collected approx. 500 images of each person. Afterward for profiling people we deployed surveillance cameras on different locations of campus. Finally, acquisitioned data has been transformed to Temporal Knowledge Graph. We will discuss data collection process details in dedicated section. Knowledge Graph is simply a graph representing interconnections between entities [2]. For analysis of dataset we have modelled some semantic queries. Some sample queries are as under:
Where were XYZ person went all the day today?
Which routes Prof. usually takes in going to department office?
How many students were present on main gate in today at XYZ time today?
How many people were present on XYZ seminar/ workshop?
How many people were present in the Research Lab when short-circuit happened.
Which places Alice visited today?
Which areas of campus faced Student congestion today?
How many students attended the particular event on specific date?
How many faculty members were present in today's meeting?
How many ambulances/ fire fighters reached on time when certain event occurred?
How many people joined in the opening ceremony of department/ Lab?
How many people visited the particular department/ departments?
What are the peak time of admission in terms of students?
How many and which departments Alice visited in her last visit.
Which library/ department Bob visits most often?
How many students visited accounts office today?
How many students visited Chancellor/ Vice Chancellor office today?
How many people were at Cafeteria at XYZ time?
How many students were present at playground today?
How many students were present at gymnasium today?
The list is not exhaustive and many questions related to current and past events are required to be probed while future can be predicted. Proposed dataset can be used in different domains like citizen profiling, workplace profiling, parental monitoring, student profiling, autonomous vehicle profiling etc.

Key words:
Temporal Knowledge Graph, Semantics, Querying, Profiling, Surveillance, Smart City

1. Introduction

This literature introduces a novel dataset for profiling following are the key highlights of dataset:

(i) Proposed dataset can be used in multiple profiling (like students, workers, managers, faculty members, citizens, vehicle etc.) tasks.
(ii) Proposed dataset can be useful for government agencies, Industrial and institutional management, parental monitoring, object detection and tracking (vehicle etc.).
(iii) Proposed data can be extended in different terms like one can enrich more relationships and temporal information from different mediums like social media. Moreover, one can introduce Multi Agent Systems for data acquisition.
(iv) Proposed dataset can be used for retrieval of temporal and semantic information about person.
(v) Proposed dataset can be used for semantic inferencing.

2. Dataset

The following table briefly shows the overview of our dataset. The mentioned data are extracted from our graph. Proposed dataset is also available in CSV file format. Here nodes represent person and locations of campus. Label represents how many constraints we have applied on dataset. Relationships shows number of relationship

Manuscript received November 5, 2019
Manuscript revised November 20, 2019

available in dataset. Lastly, relationship types as mentioned earlier is ten.

Figure 1 shows the graph model for maintaining temporal information for profile. Knowledge Graph model that is (Subject, Predicate, Object, Timestamp). As the location of the person updates the temporal information that is timestamp will also update..

Table 1 shows the representation of temporal relationships in Knowledge Graph in this work. Table 2 shows the current approaches for temporal modelling [3] and [4].

Table 1: Overview for complete dataset

Nodes	136
Labels	4
Relationships	3762 (Mostly temporal)
Relationship Types	10
Available Formats	CSV and Neo4j Graph database

Table 2: Temporal Knowledge Model

Information	**Relationship**
(Siraj, Study_at, Maju)	Study_at
(Siraj, Located_at, Maju, 2019-02-19 03:53)	(Located_at, Timestamp)

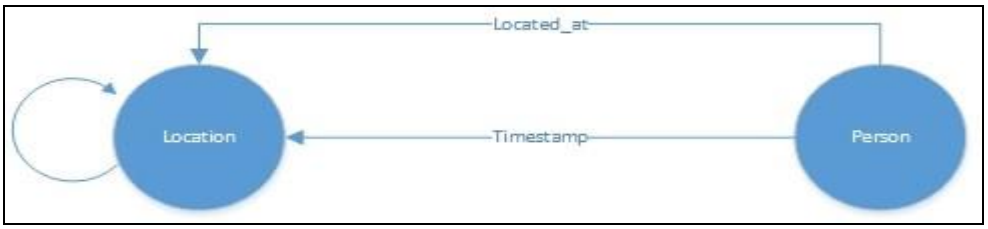

Fig. 1 Graph model for maintaining Temporal Information

Table 3 shows the tabular snapshot extracted from Knowledge Graph. Several interesting and intelligent queries have been answered in an autonomous way.

Figure 2 shows the complete Knowledge Graph extracted from Neo4j desktop.

Table 3: Tabular Snapshot of Temporal Knowledge Graph

Person	**Locations**	**Relationships**	**Timestamp**
Siraj Munir	Programming Lab	Located_at	2019-02-19 03:53
Imran Jami	Main Gate	Located_at	2019-02-19 04:53
Jahanzaib	Canopy	Located_at	2019-02-19 05:55
Reyan	Library	Located_at	2019-02-19 10:53
Mohib	Parking	Located_at	2019-02-19 02:53
Asad	Cafeteria	Located_at	2019-02-19 01:53
Shaukat Wasi	Canopy	Located_at	2019-02-19 01:50
Umair Lakahani	Cafeteria	Located_at	2019-02-19 01:30
Hamza Gaya	Library	Located_at	2019-02-19 04:53
Jahanzaib	Library	Connected with	2019-02-19 05:45

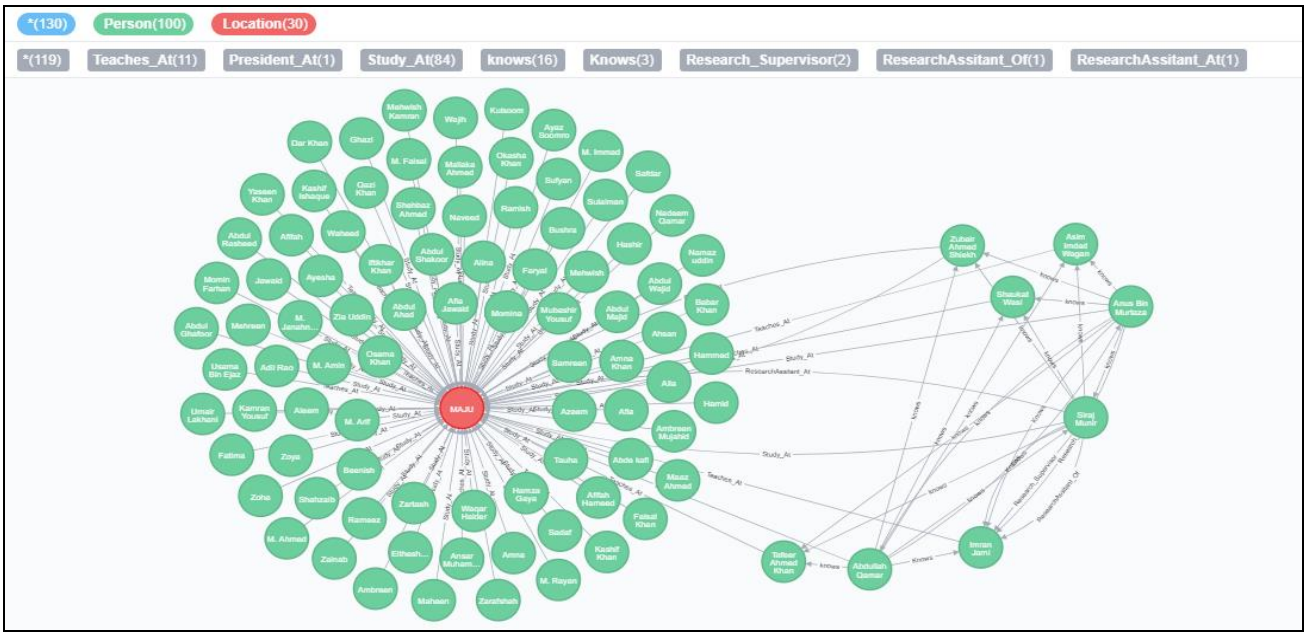

Fig. 2 Knowledge Graph representation

3. Experimental Design, Materials, and Methods

In this work we have used surveillance based mechanism for data acquisition. During whole phase of data acquisition, we have used two sources (i) RaspberryPi with PiCamera module and (ii) Surveillance feed provided collected from University campus. Figure 3 shows the workflow that is from data acquisition to Temporal Knowledge Graph. As shown in figure there are four major steps (i) Extract data from surveillance sources, (ii) Dump detected person semantic information in MongoDB,
(iii) Pre-process data and transform it into CSV and (iv) Extract Temporal Knowledge Graph from CSV using Neo4j.
Figure 4 shows the temporal Knowledge Graph. Furthermore, we can also run semantic queries on mentioned Knowledge Graph using Cypher an open source graph database supported NoSQL query language [5].

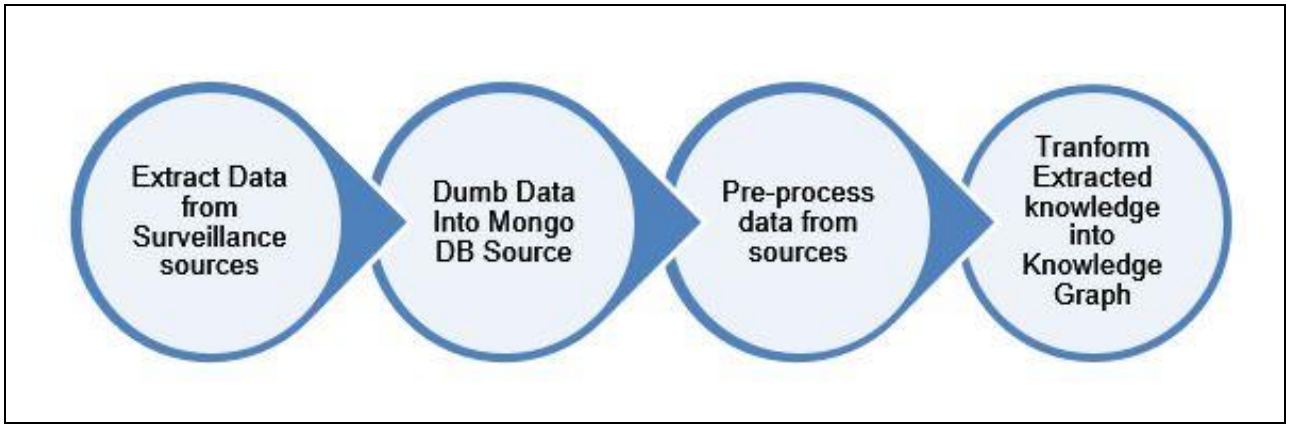

Fig. 3 Data Acquisition workflow

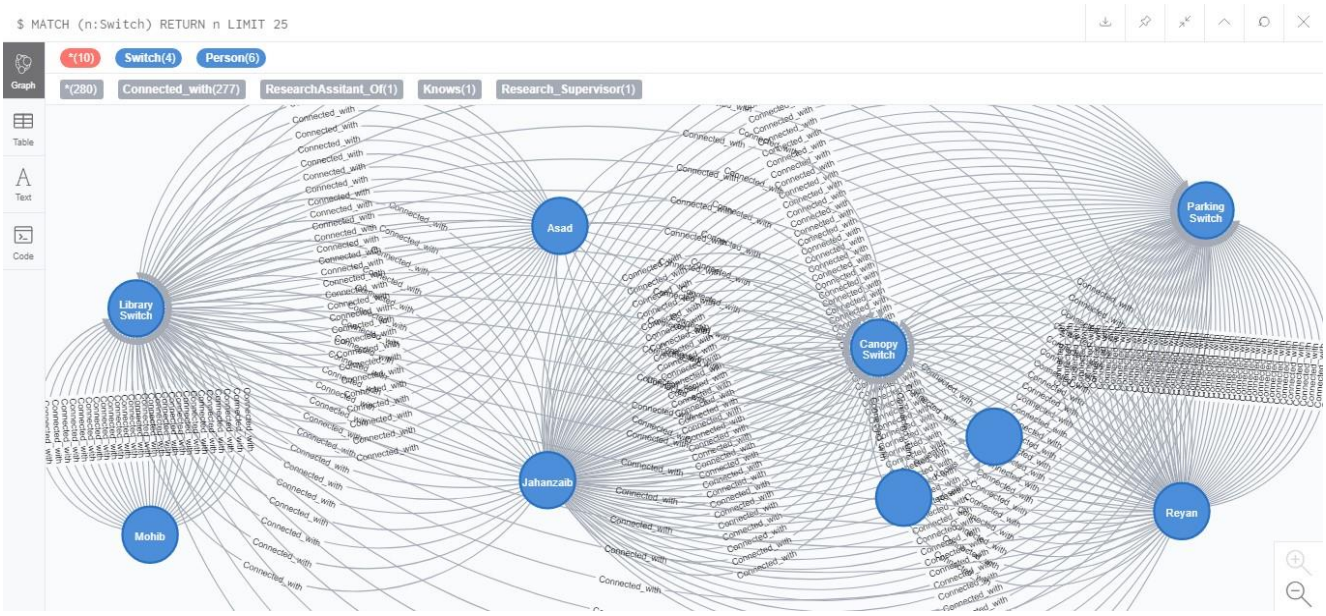

Fig. 4 Temporal Knowledge Graph representation

4. Conclusion

Knowledge Graphs are decent way to represent semantics. However, this Knowledge is gathered from multiple domains so it can be incorporated in the system for effective tracking of entity. For seamless access to the repository, Cloud based solution has been used for storage. Query language that is Cypher query language is used for information extraction. For prediction and detection of meaningful patterns inferencing has been done on tracked data. Query engine could have an intelligent agent which can take queries and retrieve related records by traversing graph. Through which we can ultimately resolve data privacy and security concerned issues of Citizen Profiling. We can also make specialized ontologies for individual purposes and on the top of them implement Knowledge Graph so that we can process intelligent queries over them effectively. Effective implementation of Citizen Profiling would also add value to industry 4.0 revolution. For instance, X company owner want to monitor his employee like managers, developers, non-technical labor etc. Through Citizen Profiling he can easy monitor his employee where was general manager when X event happened, where was foremen when machinery failed etc.

References

[1] Lawrence, S., Giles, C.L., Tsoi, A.C., and Back, A.D.: 'Face recognition: A convolutional neural-network approach', IEEE transactions on neural networks, 1997, 8, (1), pp. 98-113

[2] Paulheim, H. 'Knowledge graph refinement: A survey of approaches and evaluation methods'. Semantic web, Volume 8 Issue 3, 2017. pp. 489-508.

[3] Gottschalk, S., and Demidova, E.: 'EventKG: A Multilingual Event-Centric Temporal Knowledge Graph'. Proc. In Proceedings of the 15th Extended Semantic Web Conference (ESWC 2018) 2018 pp. Pages

[4] Garcia-Duran, A., Dumancic, S., and Niepert, M.: 'Learning Sequence Encoders for Temporal Knowledge Graph Completion', in Editor (Eds.): 'Book Learning Sequence Encoders for Temporal Knowledge Graph Completion' (Association for Computational Linguistics, 2018,), pp. 4816-4821

[5] https://neo4j.com/developer/cypher-query-language/, accessed 25-6-19 2019

Siraj Munir received the BS. in Computer Science from Indus University and M.S. in Computer Science, from Mohammad Ali Jinnah University. He is currently working as a Research Internee at Mohammed Ali Jinnah University His areas of interest include Machine Learning, Semantic Web, and Information Extraction.

Syed Imran Jami received his BS in Computer Science from the University of Karachi in 2000, MS in Computer Science from Lahore University of Management Sciences in 2004 and PhD in Computer Science from the National University of Computer & Emerging Sciences in 2011. He is one of the founding members of the Centre for Research in Ubiquitous

Computing and associated with it since 2006. He also worked with Haptics Research Lab and Pervasive and Networked Systems Research Group at Deakin University, Australia. Imran Jami has authored six journal papers and eight conference papers based on his doctoral dissertation. He is also a member of several professional organizations including ACM, IAENG, SIAM etc.

Shaukat Wasi received his MS and PhD degree in Computer Science from the National University of Computer and Emerging Sciences (FAST-NU), Karachi, Pakistan in 2006. He completed his BS (Computer Science) from the University of Karachi, Pakistan in 2003. He is a Associate Professor at Mohammed Ali Jinnah University, Karachi. His research interests include ubiquitous computing, human computer interaction, information retrieval and extraction and text mining